Joint Statistical Meetings – August 1, 2016

Improving Customer Video Service with Early Problem Detection

Cheryl Flynn

AT&T Labs Research

New York, NY

Joint work with DeDe Paul and David Poole

AT&T Labs Research

Bedminster, NJ

Most customer service strategies are reactive

Customer Experiences Problem with Service → Contacts Customer Care → Repair Technician Diagnoses/Fixes Problem

Use big data analytics to detect problems sooner to provide more information to technicians and develop proactive strategies

Reducing Multiple Repair Attempts

Customer care interactions that result in multiple repair attempts **negatively impact customer satisfaction** and result in **millions of dollars** of unnecessary costs

Can we learn a model using network data that can identify customers experiencing hard to solve video problems?

Benefits

- Alert agents to more challenging cases to help with resource allocation and time management
- Provide additional training and more guidance to technicians
- Identify preventive strategies – e.g. software updates, proactive equipment replacement

Overview

Data Collection and Pre-Processing

- Customer Care Data
- Set-top Box Data

Model Selection

- Multiple Instance Logistic Regression

Experiment Results

Note: All data is anonymized and does not contain any private customer information

Customer Care Data

Use customer care data to categorize TV customers

Data description:

- Repair appointment indicator
- 4-level classification, standardized across care centers
- Note describing care interaction

Sample Classifications

Level 1 – General Category
Trouble
Billing
Information Inquiry

Level 2 – Problem Description 1
TV Intermittent
TV Impaired
TV No Service

Level 3 – Problem Description 2
Equipment Issue
Remote Not Working
DVR Recording Issue

Level 4 – Resolution
Reset Equipment
Educated Customer
Repair Appointment

Creating Labeled Data – Service Affected Groups

Identify TV service affected cases using level codes.

Distinguish between:

1. Single Repair – single repair attempt with no repeat repair within 30 days
2. Multiple Repair – 1+ repair attempts within 30 days

Creating Labeled Data – Potential Control Groups

Standard Choice: Non care-caller

- No customer care interaction during time window or 30 days before or after

Creating Labeled Data – Potential Control Groups

~~Standard Choice: Non care-caller~~

- ~~No customer care interaction during time window or 30 days before or after~~

Unknown number of “silent sufferers”
→ exclude from training

Creating Labeled Data – Potential Control Groups

~~Standard Choice: Non care-caller~~

- ~~No customer care interaction during time window or 30 days before or after~~

Unknown number of “silent sufferers”
→ exclude from training

Proposed Alternative: Non-Service Affected

- Customer called customer service
- No repair appointment required
- No customer call-back within 30 days
- No indication of any service problem in level codes or agent notes

Set-top Box Data

Construct a set of predictor variables using performance metrics collected from STBs

Performance metrics:

- 26 variables related to receiver, tuner, and VOD functionality
- Channel change and reboot indicator
- Collected every 15-minutes when STB is active

Gathered additional STB information based on conversations with domain experts:

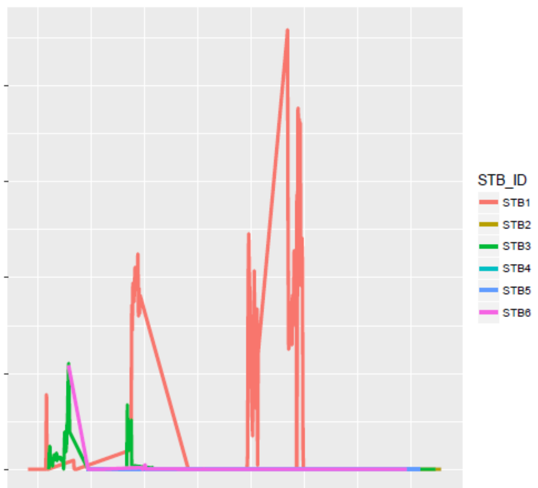
- Receiver type, vendor, and chipset type

Initial Data Observations

1. Data is spikey in nature

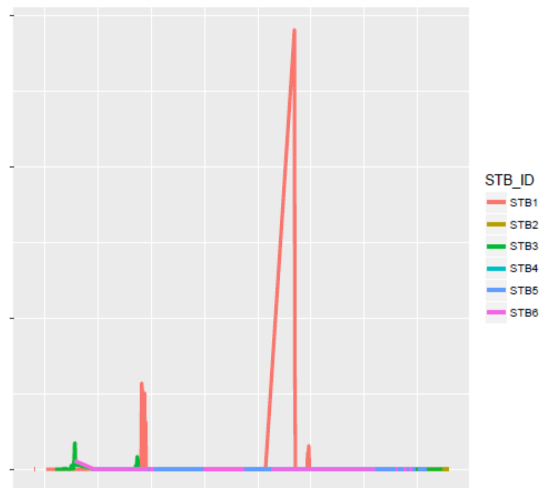
Sample TV Multiple Repair Case

Receiver Diagnostic 1



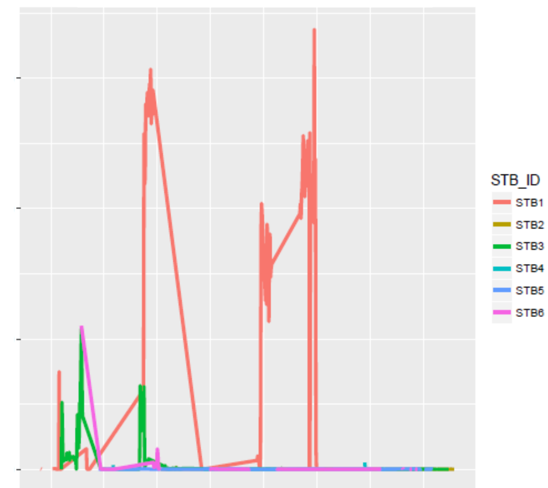
1 Week Prior to Call

Tuner Diagnostic 1



1 Week Prior to Call

Tuner Diagnostic 2



1 Week Prior to Call

2. Spikes are typically orders of magnitude higher for multiple repair cases than other categories

Idea: Use differences in maxima to separate groups

Constructing Predictor Variables

Initial data cleaning to control for sensitivity to reboot and channel change events.

Collect data over a specified time window (e.g. 7 days prior to call date) and compute the maximum value for each variable for each STB

Compare performance using raw data to performance using discretized data

Discretization

- Create a reference distribution of maxima using non-care callers (excluded from training)
- Use 5^{th}, 10^{th}, 90^{th} and 95^{th} percentiles of this distribution as cut points to create an ordinal value for training and test data

Data Set-up

Unknown STB Label

Observed Household Label

$\text{STB}_1 \rightarrow \boldsymbol{x_{i1}} = (x_{i1k})$, for $k = 1, \dots, K \rightarrow y_{i1} = ?$

$\text{STB}_2 \rightarrow \boldsymbol{x_{i2}} = (x_{i2k})$, for $k = 1, \dots, K \rightarrow y_{i2} = ?$

$\vdots \qquad \vdots \qquad \vdots$

$\text{STB}_{\text{ni}} \rightarrow \boldsymbol{x_{in_i}} = (x_{in_ik})$, for $k = 1, \dots, K \rightarrow y_{in_i} = ?$

$$y_i$$

Multiple Instance Learning

Multiple Instance Learning is a semi-supervised setting where we observe a label for a “bag” of instances but not the label for the instances themselves

Many application areas including

- Drug activity prediction (Dietterich et al., 1997)
- Image retrieval (Maron et al., 1998, and Rahmani et al., 2006)
- Sensor data (Guan et al., 2016)

Typically make a “presesence-based” assumption (Weidmann et al., 2003) where a bag is labeled positive if at least one of the instances is positive and negative otherwise

Here we consider an extension of logistic regression to multiple instance data similar to Babenko (2004), Saul et al. (2001), and Xu et al. (2004)

Logistic Regression for Multiple Instance Data

Let $\boldsymbol{H}_i = \{\boldsymbol{x}_{ij}\}_{j=1}^{n_i}$ be the collection of device information for household i with n_i devices, where $i = 1, \dots, M$.

Assume

$$p_{ij} = P(y_{ij} = 1 \mid \boldsymbol{x}_{ij}) = \frac{1}{1 + e^{-\sum_{k=1}^{K} x_{ijk}\beta_k}}$$

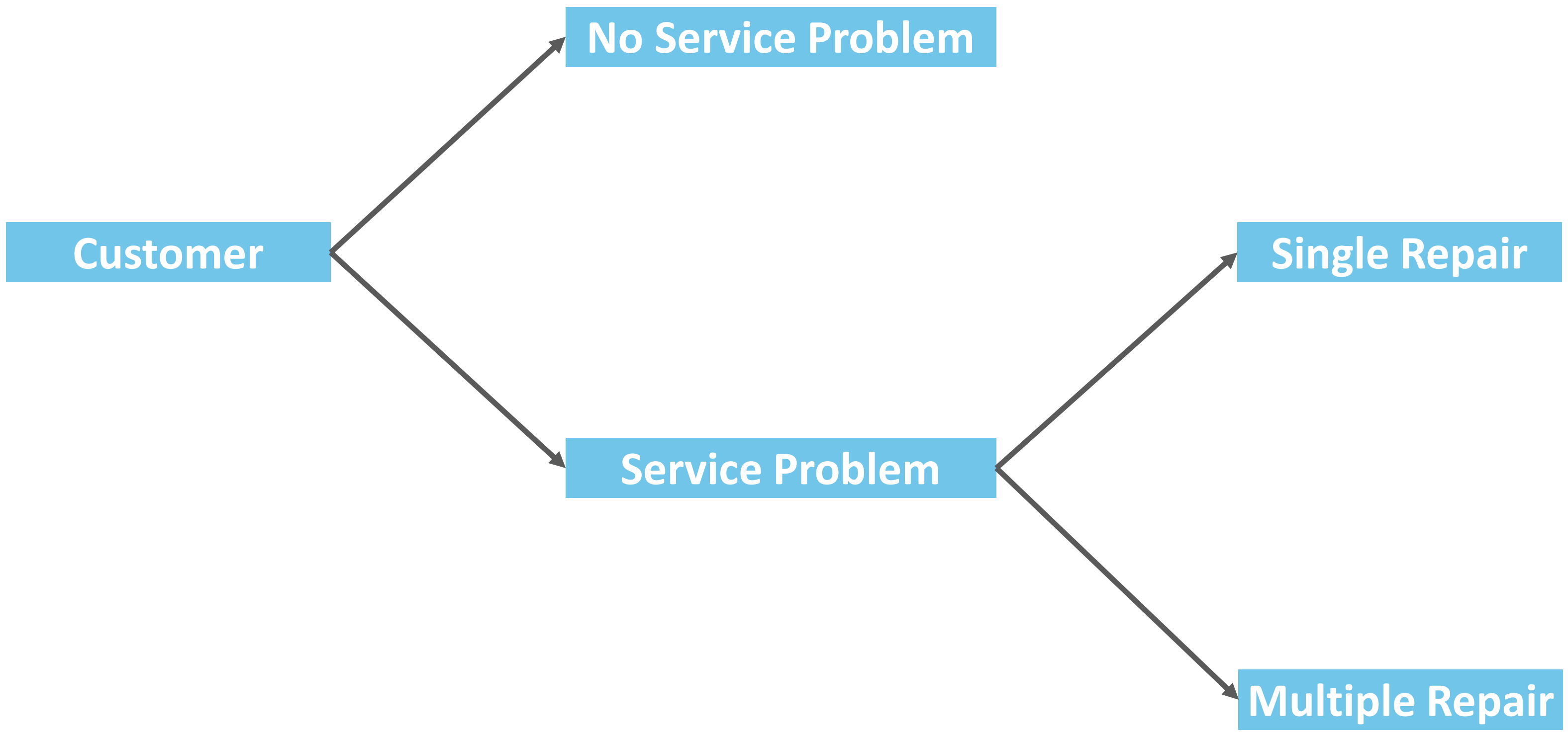
and

$$p_i = P(y_i = 1 \mid \boldsymbol{H}_i) = softmax_{1\le j\le n_i}\, p_{ij} = \frac{\sum_{j=1}^{n_i} p_{ij} e^{\alpha p_{ij}}}{\sum_{j=1}^{n_i} e^{\alpha p_{ij}}}$$

for $\alpha > 0$.

Solve for $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_k)$ using maximum likelihood estimation

Build a Hierarchical Model



$$\mathrm{P}(\text{Multiple Repair} \mid \mathrm{H}) = \mathrm{P}(\text{Multiple Repair} \mid \text{Service Problem}, \mathrm{H})\ \mathrm{P}(\text{Service Problem} \mid \mathrm{H})$$

Sample Construction

Based on publicly available industry benchmarks, we assume

- 20% of customers contact customer service
- 20% need a repair
- 25% result in a repeat repair

Construct a random sample over a two month period of 400k customers according to the above assumptions

Potential savings:

- For illustrative purposes, assume repeat repairs require an additional 1.5 repair attempts at an average cost of $100 per repair
- If there are 2k repeat repair cases a month, then this represents an annual potential savings of $3.6 million

Experiment Set-up

Time window

- Collect data for each STB over a 7 day time window
- Use first care interaction date as reference date for care callers
- Assign random reference date for non-care callers

Training/Test Sets

- Train on first month, test on second month
- Exclude non-care callers from training, but predict class for all customers in test set

Evaluation Metrics

- Cumulative capture rate (Recall) – % of multi-repair cases selected
- Hit Rate (Precision) – restricting to care-callers, % of selected cases that are multi-repair

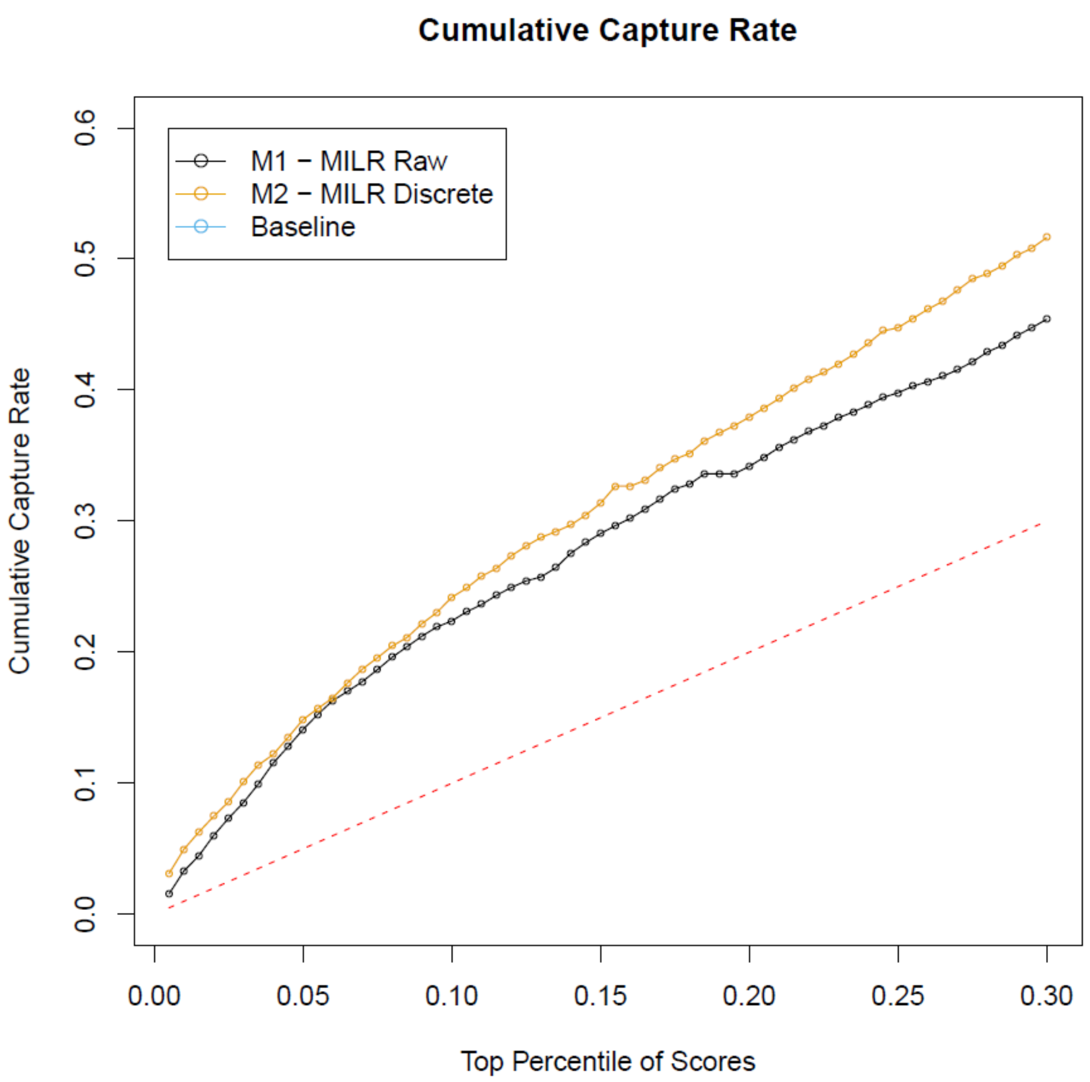
Variable Selection

Candidate Models

- Considered STB metrics, device information, and their interaction effects as potential predictors
- Fit lasso penalized logistic regression to customers with one STB
- Considered models selected along the regularization path as set of candidate models

Select final model for each MILR using $\mathrm{AIC} = -\mathrm{LL} + 2\mathrm{k}$

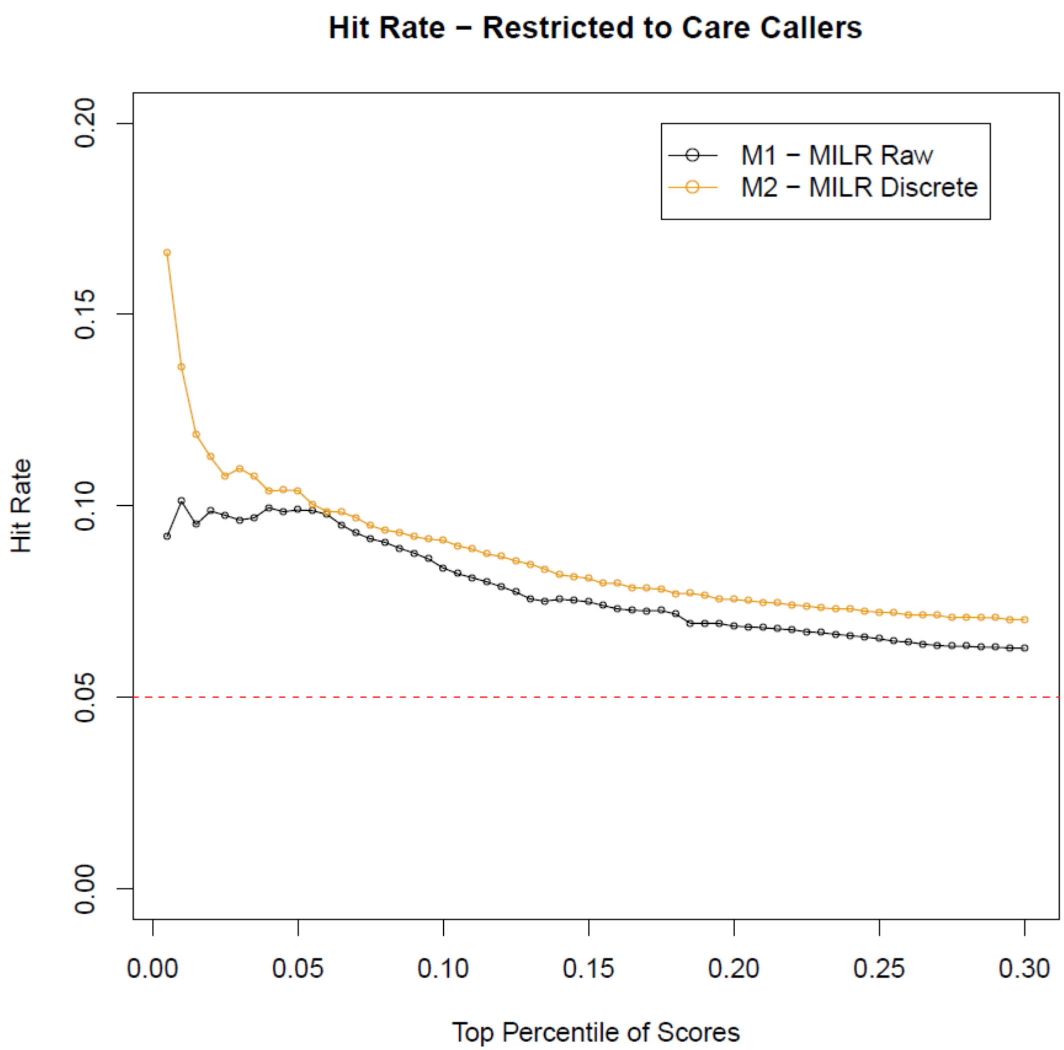
Experimental Results – Cumulative Capture Rate



- We capture about 25% of Multiple Repairs within the top 10% of scores using the discrete predictors, i.e. 2.5x the baseline rate

- Model with discrete predictors always performs at least as well as the model with raw predictors

Experimental Results – Hit Rate



- The hit rate for the MI instance model with discrete predictors is about 1.8x the baseline rate for the top 10% of scores
- Model with discrete predictors at least as well as the model with raw predictors

Summary

- Model effectively captures customers who are likely to be experience difficult to solve technical video problems that are likely to result in multiple repair attempts
- Targeting the top 10% of scores, the model captures 25% of multiple repair attempts
- Based on our example with 400k customers, this equates to $0.9 million of potential annual savings
- Precision can be improved by targeting specific customer groups

Future Work

Use information to take action that will improve the customer experience and reduce costs

- Schedule more time for the appointment
- Assign more experienced technician
- Provide better guidance to newer technicians
- Proactive intervention

Work with business leaders to develop a field trial to test effectiveness of the model and assess the potential benefits to customers and cost savings

Investigate additional approaches to multiple instance learning that specifically model correlation among STBs and time-series nature of the data

References

B. Babenko. Multiple instance learning: Algorithms and applications. Technical report, Department of Computer Science and Engineering, University of California, San Diego, 2004.

T. Dietterich, R. Lathrop, and T. Perez. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence, 89: 31-71, 1997.

X. Guan, R. Raich, and W.K. Wong. Efficient Multi-Instance Learning for Activity Recognition for Time Series Data Using an Auto-Regressive Hidden Markov Model. ICML, 2016.

O. Maron and T. Lozano-Perez. A Framework for Multiple-Instance Learning. NIPS, 1998.

R. Rahmani and S. Goldman. MISSL: Multiple-Instance Semi-Supervised Learning. ICML, 2006.

L. Saul, M. Rahim, and J. Allen. A statistical model for robust integration of narrowband cues in speech. Computer Speech and Language, 15(2): 175-194, 2001.

Weidmann, N., Frank, E., and Pfahringer, B. A two-level learning method for generalized multi-instance problems. In In Proceedings of the Fourteenth European Conference on Machine Learning, 2003.

X. Xu and E. Frank. Logistic Regression and Boosting for Labeled Bags of Instances. Proceedings Pacific Asia Conference Knowledge Discovery and Data Mining, 2004.